DOCUMENT RESUME

ED 067 092 JC 720 224

AUTHOR Calista, Donald J.
TITLE A Reassessment of College Students' Instructional Expectations and Evaluations.
PUB DATE [72]
NOTE 30p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Course Evaluation; *Evaluation Methods; Evaluation Techniques; *Expectation; *Junior Colleges; Learning Processes; Rating Scales; *Teacher Evaluation

ABSTRACT

This paper explores two aspects of student evaluations of college teaching: (1) a reformulation of end-of-term ratings, by defining them in relation to initial student expectations; and (2) it presents a research design which studies this reformulation in the natural setting of the classroom. An expectations instrument, using semantic differential scales, was administered at the beginning and end of a semester to 209 social science students at three colleges. Some findings were: (1) there were statistically significant differences between expectations and evaluations, even for those classes where the actual evaluations were quite high; (2) freshmen and sophomores at all three colleges exhibited similar expectations; (3) sophomores, rather than freshmen, consistently revealed higher evaluations; and (4) compared with sophomores, freshmen expectations appeared more in keeping with their evaluations. It was concluded that an end-of-course rating can be a generally reliable indicator of student reaction, but an accurate appraisal of the dynamics of the teaching-learning situation requires input of teacher and student expectations. (RN)

ED 067092

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

# A REASSESSMENT OF COLLEGE STUDENTS' INSTRUCTIONAL EXPECTATIONS AND EVALUATIONS

Donald J. Calista

Chairman, Behavioral and Social Sciences Department

Bennett College

New York

[1972]

JC 720 224

UNIVERSITY OF CALIF.
LOS ANGELES

NOV 1 1972

CLEARINGHOUSE FOR
JUNIOR COLLEGE
INFORMATION

ABSTRACT

A Reassessment of College Students'

Instructional Expectations and Evaluations

Donald J. Calista

Chairman, Behavioral and Social Sciences Department, Bennett College

Questioning the conventional end-of-course ratings of professors, an evaluations model was introduced to extend the notion of ratings to incorporate measures of student classroom expectations. As a working framework for classroom role structure, the study employed Stogdill's (1959) group achievement formulation to distinguish between those expectations associated with institutional dynamics--defined as situational--and those related with the classroom--defined as contextual. Test at the beginning and retest at the closing of a semester was administered to 209 students in 8 social science classes at two two-year colleges and one four-year college. Developed from previous Critical Incident and factor analysis techniques, the 20-item instruments used modified Semantic Differential scales. The major hypothesis found statistically significant differences between each class' expectations and evaluations, even for high evaluations. Year in college was not a significant intervening variable for either expectations or evaluations. If the evaluations demonstrated the direction of expectations, but failed to indicate their intensity, the study concluded that extrapolations made solely from evaluations were of questionable value in measuring the range of student-teacher interaction. Redefining the expectations-evaluations framework seemed appropriate and further research suggestions were offered.

If the literature on student ratings of college courses dates from the 1920s (Costin et al., 1971), its impact on educational practice is far from conclusive.* Despite the greater use of statistical refinements, there is little, if any, agreement as to what the ratings actually measure (Cohen and Brawer, 1969). Is it teaching or learning, or some combination--subtle or otherwise--of the two that is judged? Often researchers skirt any theoretical--even semantic--issues by not discussing whether student ratings are in fact evaluations. It seems that end-of-term ratings per se should now be extended to incorporate certain goal-setting and monitoring mechanisms. Only in this way can the traditional notion of ratings become a more inclusive concept of evaluations. This paper offers two alternatives to the present unsettled state of student evaluations of college teaching. First, it seeks to reformulate end-of-term ratings, by defining them in relation to initial student expectations. And, secondly, the paper presents a research design that employs this reformulation.

## Expectations-Evaluations Framework

If there exists substantial literature on college student expectations related to institutional dynamics, similar research directed toward college courses remains scant (Stern, 1970). Nonetheless, to differentiate between them, those expectations related to institutional press as explored by Stern (1970:6-7), may be called <u>situational</u> and those expectations directed

---

*My special thanks for generous help in completing this study are due to J. William Nystrom, Charles L. Palermo and Max Weiner.

toward the classroom, contextual. Their essential difference being that situational expectations may be more diffuse and generalized, whereas contextual expectations are related more to group objectives and role expectations. This clarification seems necessary as much of the research contoversy over the problem of student situational expectations stems from the unclear distinction made between situational and contextual categories which, at least, for etiological purposes, must be identified and considered separately (cf. Feldman and Newcomb, 1969, I:78-80). To make that distinction, contextual expectations may be considered as those role and group objectives students generate in relation to college courses and instructors. A further separation also needs to be struck with consensual expectations as discussed by Newcomb (1962). The key difference between consensual--and contextual--expectations is the former's open-endedness or self-defining qualities, especially in relation to peer group membership (Newcomb, 1962:469-475; cf. Gross et al., 1958:48-69).

Although there is considerable work on contextual expectations in small-group research (Cartwright and Zander, 1968; Berger et al., 1969), there is little understanding of how they relate to college classroom dynamics. The most suggestive formulation for this set of contextual expectations has been provided by Stogdill (1959) in his general work on group expectations. Stogdill defines expectation as a "readiness for reinforcement" which "is a function of drive, the estimated probability of occurrence of possible outcome, and the estimated desirability of outcome" (1959:62). By drive he means "the level of tension reactivity exhibited by an organism," which would become activated into readiness for reinforcement defined as "the extent to which an individual is prepared or unprepared to experience, or reconciled to the prospect of

experiencing, a possible outcome." In turn, actual outcomes would be mediated by an individual's prediction, judgment, or guess relative to the likelihood that a given event will occur," as well as by the estimated desirability which he defines as "an individual's judgment relative to the satisfyingness of, need for, demand for, appropriateness of, or unpleasantness of, a possible outcome" (1959:62). In the classroom, these two estimates determine student expectations associated with projected outcomes of teacher behaviors. Aware that some question might be raised regarding the positive or negative valence of the estimates, Stogdill adds that the "estimate of probability and estimate of desirability are not opposite ends of the same continuum...." Instead, these estimates "interact to determine the level of expectations." And, most relevant to the classroom, "this interaction is formulated in terms of what is uncertain in the future as well as in relation to what has been learned in the past" (1959:128).

If this concept of expectations can be understood as an interaction of probable and desirable estimates, the particular behaviors are dependent upon the classroom role structure. Accordingly, Stogdill first distinguishes between the place of the formal structure of groups, making it possible to determine the different expectations actually associated with individual instructors by students. As such, Stogdill discriminates between expectations "attached to position rather than the occupant of a position." Since the classroom's formal structure prescribes the professor's status and function which "are defined for a position rather than for any given occupant of the position," Stogdill emphasizes the relevance of expectations related to role structure--and this is most appropriate here (1959:123). For, if the ideal professor studies--which

complement the instructor ratings research--have identified distinct classes of role behavior states associated by students with instructional expectations (Perry, 1969; Lewis, 1968), then, examining the interplay of these traits seems to be the next logical research step.

Consequently, expectations as outlined by Stogdill have been used in this paper as the role-related set of probable and desirable estimates of instructor behaviors for which students exhibit a readiness of reinforcement. This formulation permits classroom role expectations--and subsequent evaluations--to be understood as both a function of the normative aspects of a classroom coupled with individual instructor behavior characteristics. The interaction of student estimates of desirable and probable instructional behavior outcomes, in other words, not only determines evaluations, as the ratings or ideal professor studies suggest, but it also activates student expectations (cf. Berger et al., 1969). Simply, the concept of expectations employed here includes measuring the readiness for projected outcomes.

By contrast, it seems that the model adopted by typical end-of-course ratings studies actually could impede an understanding of student expectations in relation to their evaluations. By excluding measures of probability and desirability these studies neglect to incorporate the expectations that ordinarily precede evaluations. Nor can expectations be assumed to be inherent in evaluations, for this has not been empirically established at all. According to the definition of expectations used here, for example, an affirmative or negative end-of-term rating in any instructional behavior category could have exceptional meaning only if initial expectation inputs are taken into account. In the ratings studies, however, there seems to be no way to ascertain whether students expect a

teacher to be friendly, even though he, in fact, turns out to be quite friendly. The student or class could be, nonetheless, expecting to achieve more organized knowledge about subject matter for which the students rate this less affirmatively. The degree to which either sheds light on actual student expectations may be limited. In the ratings studies--and more so in practice--the tendency would be to total the two results, perhaps, one equalizing or cancelling the other. Extrapolations of student expectations made from their evaluations could be, in short, of questionable value.

## Research Design

Even when two recent researchers attempted to recast the typical end-of-course ratings design, both stuck to surveying college students about their expectations at the completion of the course (Phillips, 1964; Twa, 1970). In keeping with the reformulation of the expectations-evaluations framework now being proposed, this design was longitudinal and multi-institutional. Following Gage (1961;17), it further proposed relating "changes in student/s/--the dependent variable--to some measure of the teacher's behavior or classroom experience." An expectations instrument, later modified for evaluations, was administered at the start of a semester to 209 students about equally divided into 8 introductory social science classes in 3 colleges. Students were not informed that the instrument would be retaken for evaluations purposes at the end of the semester. The three colleges were chosen because each shares the explicit institutional goal of stressing the teaching function. Alpha is a well-established two-year college for women enrolling about 400 students; Beta a newer comprehensive community college enrolling about 2,500 students; and Gamma is a 100 year old co-educational four-year college enrolling

about 1,600 students. The colleges are located within a 100 mile radius of New York City.

The instrument employed developed principally from two different analytical ratings techniques. One of the most widely used techniques can be found in the University of Michigan studies (Isaacson, 1963; McKeachie, 1963; McKeachie, 1969), which applied factor analysis to items gathered from previously used ratings instruments. From such a pool of 145 items, these studies derived 34 significant items (Isaacson et al., 1963:345). The other research technique identified items through the Critical Incident method (Ryans, 1960:79-83). From structured and unstructured procedures Owen identified an inventory of critical incidents students associate with college instructors' behaviors. His study further demonstrated a high degree of intercorrelation between its Critical Incident items and the Michigan ones (1967:140). Thus, the items for the present study's expectations-evaluations instrument were selected both from Owen's study and from Michigan's pool. In all, a total of 20 items were used for the instrument. These items were, in turn, clustered within the six Critical Incident instructional behavior categories (1967:47).[1] Because both the items and the category clusters were identified from estimates of desirable and probable behaviors associated with instructional roles, they can serve as a working approximation of contextual expectations.

By comparison, if the Critical Incident and factor analysis techniques provided highly correlated instruments items, there is little consistency

---

[1]The six instructional behavior categories employed were: I-Presentation of Material-Content, Structure, and Scope; II-Presentation of Material-Student Participation; III-Presentation of Material-Instructor's Style; IV-Teacher and Student Rapport-Class Interaction; V-Evaluation of Students; and, VI-Requirements of Students.

in scaling techniques for end-of-course ratings research. Though variety in scaling practice can be advantageous, there has been no substantial attempt on the part of ratings' researchers to integrate their work with previous ones (Costin et al., 1971). For example, if one study adopted a nine-point sentence completion technique with the opposite poles being some variation of: "I found the textbook in this course to be: 1) Very dull" to "9) Very interesting," (McClelland, 1970:89) another used adjectives such as "probing" or "sympathetic" on an open-ended seven point scale (Turner et al., 1969:67). A need for more uniform scaling technique seemed obvious.

One scaling technique already employed by two recent investigators, Lewis (1968) and Rees (1969), was the Osgood Semantic Differential (1957). In employing a seven-point scale, the semantic differential permitted greater amplitude at the extremes, while incorporating a mid-point as well. The substantive criticism raised concerning its scale had been the usual one regarding the neutrality of the mid-point (Snider and Osgood, 1969). Yet it appeared, from recent evidence, that this did not impair the usefulness of the technique (Oetting, 1967). Despite their somewhat different designs, the Lewis and Rees studies suggested promising applications of the semantic differential to evaluations research. Lewis used the bipolar adjectives, for example, to determine whether college professors were perceived in uniformly sterotypic terms, and found that different disciplines evoked significantly divergent images in students (1968:189). Unlike these studies, however, the current study employed only one bipolar pair of adjectives per instrument item. For instance, "What do you expect the instructor's knowledge of the subject to be"?

Superior___: ___: ___: ___: ___: ___: ___: ___Inferior

In addition, in order to determine the appropriateness of the expectations-evaluations instrument the last item of each recorded student appraisal of the instrument. This, measure, usually absent from ratings instruments, read: How do you rate the ability of this survey to determine your own expectations [evaluations] for this course?

Strong___: ___: ___: ___: ___: ___: ___: ___Weak

Though this was certainly not an internal control, it would still provide some indication of the effectiveness of the Semantic Differential.

## Hypotheses Considered.

The major hypothesis would naturally compare the expectations with the subsequent evaluations. Stated in the null form: One, within any class there would be no statistically significant difference in student expectations in the six Critical Incident categories and the subsequent student evaluations. Because year in college is so prominent in research related to institutional press (situational expectations), three hypotheses were raised regarding college year. Two, freshmen would not exhibit statistically significant higher expectations compared with other classmen in any one of the six Critical Incident categories; three, freshmen would not exhibit statistically significant higher evaluations compared with other classmen in any one of the six Critical Incident categories; and, four, there would be no statistically significant difference between freshmen expectations and their evaluations compared with other classmen in any one of the six Critical Incident categories.

The question of year in college relates specifically to the notion of the "Freshman Myth" (Stern, 1970:173). In his studies dealing with situ-

ational expectations, he found that freshmen "look forward to high levels of activities relevant to both academic and non-academic press...." Yet, "this does not correspond to the actual characteristics of [their] schools at all? (1970:92). Even though these studies included general questions about faculty instruction and courses, how close the resulting conclusions would be when contextual expectations were differentiated from the situational ones remained an open question. Hence, the three hypotheses involving year in college.

## Findings and Discussion

Since the data will be presented in terms of the Critical Incident instructional categories, they will be abbreviated as follows: I--Content; II--Participation; III--Style; IV--Rapport, V--Evaluation; and, VI--Requirements. It should also be mentioned that Alpha's classes are cited as A, B, C and D; Beta's as E and F; and, Gamma's as G and H. For each of the hypotheses, two-tailed tests were used because the null form did not include a prediction of direction. Statistical significance was minimumly accepted at the .05 level of judgment.

The first--and major--hypothesis considered the changes in each sample class. To test the null hypothesis of no differences between expectations and evaluations for each Critical Incident category, group means were compared using the $t$ test between correlated means (Mueller et al., 1970:417). In each case degrees of freedom were determined by the number of test pairs minus one. The relevant data can be found in Tables 1, 2 and 3. In Beta and Gamma colleges significant differences were found in virtually all of the categories in each of the classes and college totals. Most of the differences were at the .01 level. Alpha college classes reported some significant differences in three of the four classes. The college total

TABLE 1.--Means and Standard Deviations for Expectations and Evaluations Compared by CI Category for the four classes in Alpha College and College total

| Class and College | | CI Category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Content | | Participation | | Style | | Rapport | | Evaluation | | Requirements | | |
| | | $\overline{X}$ | SD | $\overline{X}$ | SD | $\overline{X}$ | SD | $\overline{X}$ | SD | $\overline{X}$ | SD | $\overline{X}$ | SD | (N) |
| A | Ex. | 26.1 | .8 | 17.5 | 2.4 | 19.5 | .3 | 19.6 | .9 | 16.2 | 2.3 | 18.8 | 1.0 | (23) |
| | Ev. | 25.0 | 1.6 | 17.9[a] | 1.7 | 18.8 | 1.8 | 19.2 | 1.4 | 16.9 | 2.6 | 17.5[a] | 1.5 | |
| B | Ex. | 24.8 | 2.2 | 16.5 | 2.5 | 18.0 | 2.5 | 19.4 | 1.1 | 16.2 | 2.7 | 17.9 | 1.6 | (21) |
| | Ev. | 25.6 | 1.5 | 16.3 | 1.7 | 18.8 | 1.2 | 18.7 | 1.6 | 17.7 | 1.9 | 18.1 | 1.8 | |
| C | Ex. | 25.2 | 1.5 | 17.0 | 1.6 | 18.7 | 1.3 | 19.7 | .7 | 16.7 | 2.9 | 18.2 | 1.5 | (29) |
| | Ev. | 24.3[a] | 1.4 | 15.5[a] | 2.9 | 18.6 | 1.6 | 18.9 | 1.6 | 16.5 | 2.3 | 17.1[a] | 2.1 | |
| D | Ex. | 24.6 | 2.2 | 17.3 | 1.6 | 19.0[b] | 1.5 | 19.9[b] | .9 | 17.0 | 1.7 | 17.9[a] | 2.2 | (23) |
| | Ev. | 23.0 | 3.6 | 14.2[b] | 3.3 | 16.5[b] | 3.7 | 17.9[b] | 2.1 | 16.5 | 1.9 | 16.6[a] | 1.8 | |
| Alpha Total | | | | | | | | | | | | | | |
| | Ex. | 25.2[a] | 1.8 | 17.1 | 2.1 | 18.8 | 1.6 | 19.7 | .9 | 16.6 | 2.5 | 18.2[a] | 1.9 | (96) |
| | Ev. | 24.4[a] | 2.4 | 15.9[b] | 2.9 | 18.2 | 2.4 | 18.7[a] | 1.8 | 16.8 | 2.3 | 17.3[a] | 1.9 | |

[a] $t$ ratios significant at .05 level.

[b] $t$ ratios significant at .01 level.

TABLE 2.--Means and Standard Deviations for Expectations and Evaluations Compared by CI Category for the two classes in Beta College and College total

| Class and College | | CI Category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Content | | Participation | | Style | | Rapport | | Evaluation | | Requirements | | |
| | | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | (N) |
| E | Ex. | 23.4 | 2.6 | 15.3 | 2.9 | 17.8[b] | 2.2 | 19.2[b] | 2.0 | 14.4[b] | 2.8 | 16.2[b] | 2.5 | (25) |
| | Ev. | 18.6[b] | 4.5 | 11.3[b] | 3.5 | 11.0 | 4.5 | 11.7 | 4.2 | 10.5 | 3.8 | 12.6[b] | 3.6 | |
| F | Ex. | 22.9[b] | 2.4 | 15.5[b] | 2.0 | 17.9[b] | .8 | 18.5[b] | 1.2 | 16.2[b] | .7 | 16.8[b] | 1 5 | (24) |
| | Ev. | 18.6 | 3.2 | 13.0[b] | 2.3 | 14.5[b] | 1.4 | 15.5[b] | 1.3 | 12.8 | 2.2 | 13.0[b] | 2.4 | |
| Total | | | | | | | | | | | | | | |
| | Ex. | 23.2[b] | 2.5 | 15.4 | 2.5 | 17.9[b] | 1.6 | 18.9[b] | 1.7 | 15.3[b] | 2.3 | 16.5[b] | 2.1 | (49) |
| | Ev. | 18.6[b] | 3.9 | 12.1[b] | 3.1 | 13.2[b] | 3.6 | 13.5[b] | 3.7 | 11.6 | 3.3 | 12.8[b] | 3.1 | |

[a] t ratios significant at .05 level.

[b] t ratios significant at .01 level.

TABLE 3.--Means and Standard Deviations for Expectations and Evaluations Compared by CI Category for the two classes in Gamma College and College total

| Class and College | | CI Category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Content | | Participation | | Style | | Rapport | | Evaluation | | Requirements | | |
| | | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | (N) |
| G | Ex. | 24.9 | 1.8 | 16.3[a] | 2.2 | 18.1 | 2.3 | 19.5[b] | 1.9 | 15.1[a] | 2.0 | 17.6[b] | 1.7 | (40) |
| | Ev. | 23.9[a] | 2.4 | 15.3[a] | 2.6 | 17.7 | 2.5 | 18.6[b] | 1.7 | 13.8[a] | 2.4 | 16.4[b] | 2.9 | |
| H | Ex. | 24.0 | 1.3 | 14.6[b] | 3.4 | 15.9[b] | 3.2 | 18.2[b] | 2.5 | 15.2 | 2.7 | 17.6[b] | 2.2 | (24) |
| | Ev. | 21.0[b] | 3.0 | 12.1[b] | 2.9 | 13.2[b] | 3.8 | 13.6[b] | 3.0 | 13.5 | 2.8 | 14.4[b] | 2.3 | |
| Gamma Total | | | | | | | | | | | | | | |
| | Ex. | 24.6[b] | 1.7 | 15.7[b] | 2.8 | 17.3[b] | 2.9 | 19.0[b] | 2.2 | 15.1[b] | 2.3 | 17.6[b] | 1.9 | (64) |
| | Ev. | 22.8[b] | 3.0 | 14.1[b] | 3.1 | 16.0[b] | 3.7 | 16.7[b] | 3.7 | 13.7[b] | 2.6 | 15.6[b] | 2.9 | |

[a] $t$ ratios significant at .05 level.

[b] $t$ ratios significant at .01 level.

produced significant differences in four of six Critical Incident categories. In all, 30 out of 48 expectations-evaluations Critical Incident comparisons reached acceptable levels of statistical significance. In only 4 instances in classes A and B did the mean differences for the evaluations outweigh the expectations; of these, one instance produced a significant difference in the Participation category of class A. Thus, the null hypothesis was rejected for the alternative that differences were to be found between student contextual expectations and their evaluations.

An analysis of the ranks for the expectations and evaluations categories confirmed tha alternative hypothesis, too. This analysis might help answer questions raised regarding the reliability of the study population. Data reported in Table 4 indicated the rank order (rho) computations for the Critical Incident expectations scales in each of the colleges. The rank differences between the colleges were negligible. Furthermore, the relative constancy of the students' expectations can be seen by comparing the rank order (rho) differences of their expectations and their subsequent evaluations. Table 5 reported this data. Despite the varying group mean differences actually recorded by each class, the students ranked the six Critical Incident categories in a rather consistent fashion. This seemed remarkable even if one were faced with the fact that the students' expectations were met with uneven outcomes. The rho ratios for the Critical Incident categories for Alpha college was .93, for Beta .76, and for Gamma .94. Certainly this would serve as a confirmation of the substantive hypothesis that expectations influence the way students record their evaluations of instructors.

TABLE 4.--Comparison of Rank Differences by CI Expectations Category for Alpha, Beta and Gamma Colleges

| CI Category | College, Rank and Mean | | | | | |
|---|---|---|---|---|---|---|
| | Alpha[a] | | Beta[b] | | Gamma | |
| | Ex. | | Ex. | | Ex. | |
| | Rank | $\bar{X}$ | Rank | $\bar{X}$ | Rank | $\bar{X}$ |
| Content | 2 | (6.3) | 3 | (5.8) | 2 | (6.2) |
| Participation | 5 | (5.7) | 6 | (5.0) | 5 | (5.2) |
| Style | 3 | (6.2) | 2 | (5.9) | 3 | (5.8) |
| Rapport | 1 | (6.5) | 1 | (6.3) | 1 | (6.3) |
| Evaluation | 6 | (5.5) | 5 | (5.1) | 6 | (6.0) |
| Requirement | 4 | (6.0) | 4 | (5.5) | 4 | (5.7) |

[a]Rho=1.0 with Gamma College.

[b]Rho=.90 with Alpha and Gamma Colleges.

TABLE 5.--Comparison of Rank Differences by CI Expectations Category with CI Evaluations Category for Alpha, Beta and Gamma Colleges

| CI Category | College, Rank and Mean | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Alpha[a] | | | | Beta[b] | | | | Gamma[c] | | | |
| | Ex. | | Ev. | | Ex. | | Ev. | | Ex. | | Ev. | |
| | Rank | $\bar{X}$ | Rank | $\bar{X}$ | Rank | $\bar{X}$ | Rank | $\bar{X}$ | Rank | $\bar{X}$ | Rank | $\bar{X}$ |
| Content | 2 | (6.3) | 2.5 | (6.1) | 3 | (5.8) | 1 | (4.6) | 2 | (6.2) | 1 | (5.7) |
| Participation | 5 | (5.7) | 6 | (5.3) | 6 | (5.0) | 5 | (4.0) | 5 | (5.2) | 5 | (4.7) |
| Style | 3 | (6.2) | 2.5 | (6.1) | 2 | (5.9) | 3 | ((4.4) | 3 | (5.8) | 3 | (5.8) |
| Rapport | 1 | (6.5) | 1 | (6.2) | 1 | (6.3) | 2 | (4.5) | 1 | (6.3) | 2 | (5.5) |
| Evaluation | 6 | (5.5) | 5 | (5.6) | 5 | (5.1) | 6 | (3.8) | 6 | (5.0) | 6 | (4.6) |
| Requirement | 4 | (6.0) | 4 | (6.0) | 4 | (5.5) | 4 | (4.2) | 4 | (5.7) | 4 | (5.2) |

[a]Rho=.93.

[b]Rho=.76.

[c]Rho=.94.

This conclusion seemed to be supported by noting student reactions to the expectations and evaluations instruments. This data--gathered from responses regarding the efficacy of the instruments--can be found in Table 6 by sex, with a t test for correlated means computed for differences. The only difference actually noted did indicate that the evaluations instrument proved to be a better indicator of student responses. Again, despite the significant differences between actual expectations and evaluations, student response to the study instruments remained relatively constant.

To test the second hypothesis that there would be no difference between freshmen and other classmen, group means for each Critical Incident expectations category were computed and reported in Table 7. Because there were so few third and fourth year students in each introductory social science course, the data were grouped by freshmen and sophomores. The t test for uncorrelated--or independent-means--(Mueller et al., 1970:407) was used for determining statistically significant differences. Degrees of freedom were determined by summing up the two test samples minus two. Since no differences were noted, the null hypothesis was therefore accepted, which in this case, served as the substantive one. Clearly, the freshmen and sophomores reported similar contextual expectations. Even if one discounts the existence of an "upper-classmen" effect for the two-year college students, no differences were reported by the four-year college sophomores.

The third hypothesis that there would be no differences in freshman and sophomore evaluations was similarly tested and cited in Table 8. As can be readily seen, the only differences noted were for sophomores--and these were significantly greater. Although sophomore evaluations were

TABLE 6.--Means and Standard Deviations of Student Responses to Part II., Item 9 on the Evaluations Instruments for Alpha, Beta and Gamma Colleges by Sex

| College | Female | | | Male | | |
|---|---|---|---|---|---|---|
| | $\overline{X}$ | SD | (N) | $\overline{X}$ | SD | (N) |
| Alpha | | | | | | |
| Ex. | 5.49[a] | .94 | (96) | -- | -- | -- |
| Ev. | 5.88[a] | .64 | | -- | -- | -- |
| Beta | | | | | | |
| Ex. | 5.40 | .76 | (25) | 5.46 | .72 | (24) |
| Ev. | 5.48 | .77 | | 5.50 | .83 | |
| Gamma | | | | | | |
| Ex. | 5.82 | .65 | (45) | 5.84 | .60 | (19) |
| Ev. | 5.87 | .51 | | 5.79 | .54 | |

[a] t ratios significant at .05 level.

TABLE 7.--Mean and Standard Deviation Comparisons of Freshmen with Sophomores by CI Expectation Category within Alpha, Beta, and Gamma Colleges

| College and Year | CI Category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Content | | Participation | | Style | | Rapport | | Evaluation | | Requirements | | |
| | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | (N) |
| Alpha | | | | | | | | | | | | | |
| Fresh. | 25.0 | 2.0 | 17.4 | 1.9 | 19.2 | 1.2 | 19.7 | .7 | 16.9 | 2.6 | 18.2 | 2.0 | (44) |
| Soph. | 25.4 | 1.6 | 16.7 | 2.2 | 18.5 | 1.9 | 19.6 | 1.0 | 16.2 | 2.3 | 18.3 | 1.7 | (52) |
| Beta | | | | | | | | | | | | | |
| Fresh. | 23.1 | 2.3 | 15.4 | 2.8 | 17.7 | 2.0 | 18.7 | 2.0 | 15.1 | 2.3 | 16.3 | 2.3 | (30) |
| Soph. | 23.2 | 2.7 | 15.5 | 2.0 | 18.2 | 1.0 | 19.2 | 1.1 | 15.6 | 2.1 | 16.7 | 1.7 | (19) |
| Gamma | | | | | | | | | | | | | |
| Fresh. | 24.3 | 1.5 | 15.5 | 3.0 | 17.2 | 3.0 | 19.0 | 1.7 | 15.5 | 2.0 | 17.4 | 2.3 | (29) |
| Soph. | 24.8 | 1.9 | 15.8 | 2.7 | 17.4 | 2.8 | 19.0 | 2.6 | 15.2 | 2.6 | 17.7 | 1.4 | (35) |

TABLE 8.--Mean and Standard Deviation Comparisons of Freshmen with Sophomores by CI Evaluations Category within Alpha, Beta, and Gamma Colleges

| College and Year | CI Category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Content | | Participation | | Style | | Rapport | | Evaluation | | Requirements | | |
| | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | (N) |
| Alpha | | | | | | | | | | | | | |
| Fresh. | 23.9[a] | 2.7 | 15.6 | 3.0 | 17.8 | 2.8 | 18.4 | 2.5 | 16.8 | 2.5 | 17.3 | 2.0 | (44) |
| Soph. | 25.0[a] | 1.9 | 16.1 | 2.7 | 18.5 | 2.0 | 19.0 | 1.5 | 16.9 | 2.0 | 17.3 | 1.9 | (52) |
| Beta | | | | | | | | | | | | | |
| Fresh. | 18.3[b] | 3.9 | 11.6[b] | 2.6 | 12.6[c] | 3.4 | 12.8[c] | 3.3 | 11.4[a] | 3.3 | 12.1[c] | 3.4 | (30) |
| Soph. | 19.2[b] | 3.8 | 13.0[b] | 3.7 | 14.1[c] | 3.6 | 14.6[c] | 3.9 | 11.9[a] | 3.3 | 13.9[c] | 2.2 | (19) |
| Gamma | | | | | | | | | | | | | |
| Fresh. | 22.6 | 2.6 | 13.8 | 2.8 | 14.7 | 4.3[b] | 15.8[a] | 3.6 | 13.7 | 2.5 | 15.4 | 2.7 | (29) |
| Soph. | 23.0 | 3.3 | 14.3 | 3.2 | 17.1 | 2.7[b] | 17.4[a] | 3.0 | 13.7 | 2.6 | 15.7 | 3.0 | (35) |

[a] t ratios significant at .05 level.

[b] t ratios significant at .01 level.

[c] t ratios significant at .001 level.

significantly higher in these nine instances, in all but one, their evaluations remained higher than freshmen. Here again, the null hypothesis was accepted, for, when differences appeared sophomore evaluations --not freshman--produced them.

Hypothesis four tested the proposition that there would be no significant differences between freshmen expectations and evaluations as compared with sophomores. The group means for this comparison by college can be found in Table 9. Here the t test for correlated means was employed. Since there were significant differences to the .001 level for both freshman and sophomore evaluations the null hypothesis became untenable. This would confirm the substantive hypothesis of this study--students record their classroom experiences in similar ways. Taking Beta college, for example, it did not make much difference whether freshmen had lower expectations than sophomores, for the actual experiences produced significantly lower evaluations for both groups. To sum up, the three hypotheses regarding college year, both freshmen and sophomores recorded expectations and evaluations that seemed to relate each other's reality rather than their year in college as such.

## Conclusions

As a study of student contextual expectations in relation to their evaluations in the natural setting of the classroom, the findings supported the substantive hypotheses, namely, that end-of-course ratings offered only a general understanding of expectations. This was affirmed in the analyses of each class and college group means as well as rank order comparisons for the Critical Incident expectations and evaluations categories. More importantly, the findings demonstrated that statistically significant differences emerged between expectations and evaluations, even

TABLE 9.--Means and Standard Deviations Compared by CI Expectations and Evaluations Category for Year in Alpha, Beta and Gamma Colleges

| College and Year | | CI Category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Content | | Participation | | Style | | Rapport | | Evaluation | | Requirements | | |
| | | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD | (N) |
| Alpha | | | | | | | | | | | | | | |
| Fr. | Ex. | 25.0 | 2.0 | 17.4 | 1.9 | 19.2 | 1.2 | 19.7[b] | 2.6 | 16.8 | 2.6 | 18.2 | 2.0 | (44) |
| | Ev. | 23.8[a] | 2.7 | 15.8[b] | 3.0 | 17.8[a] | 2.8 | 18.4 | 2.5 | 16.8 | 2.5 | 17.3[a] | 2.0 | |
| So. | Ex. | 25.4 | 1.6 | 16.7 | 2.2 | 18.5 | 1.9 | 19.6 | 1.0 | 16.2 | 2.3 | 18.3[a] | 1.7 | (52) |
| | Ev. | 25.0 | 1.9 | 16.1 | 2.7 | 18.5 | 2.0 | 19.0 | 1.5 | 16.9 | 2.0 | 17.3[a] | 1.9 | |
| Beta | | | | | | | | | | | | | | |
| Fr. | Ex. | 23.1 | 2.3 | 15.4 | 2.8 | 17.7 | 2.0 | 18.7[c] | 2.0 | 15.1 | 2.3 | 16.3 | 2.3 | (30) |
| | Ev. | 18.3[c] | 3.9 | 11.6[c] | 2.6 | 12.6[c] | 3.4 | 12.8[c] | 3.3 | 11.4[c] | 3.3 | 12.1[c] | 2.2 | |
| So. | Ex. | 23.3 | 2.7 | 15.5 | 2.0 | 18.2 | 1.0 | 19.2 | 1.1 | 15.6 | 2.1 | 16.7[c] | 1.7 | (19) |
| | Ev. | 19.2[c] | 3.8 | 13.0[a] | 3.7 | 14.1[c] | 3.6 | 14.6[c] | 3.9 | 11.9[c] | 3.3 | 13.9[c] | 3.4 | |
| Gamma | | | | | | | | | | | | | | |
| Fr. | Ex. | 24.3[b] | 1.5 | 15.5 | 3.0 | 17.2[b] | 3.0 | 19.0 | 1.7 | 15.5 | 2.0 | 17.4 | 2.3 | (29) |
| | Ev. | 22.6 | 2.6 | 13.8[a] | 2.8 | 14.7 | 4.3 | 15.8[c] | 3.6 | 13.7[a] | 2.6 | 15.4[c] | 2.7 | |
| So. | Ex. | 24.8[b] | 1.9 | 15.8[b] | 2.7 | 17.4 | 2.8 | 19.0[b] | 2.6 | 15.2 | 2.5 | 17.7 | 1.4 | (35) |
| | Ev. | 23.0 | 3.3 | 14.3[b] | 3.2 | 17.1 | 2.7 | 17.4 | 3.0 | 13.7[b] | 2.6 | 15.7[b] | 3.0 | |

[a] $t$ ratios significant at .05 level.

[b] $t$ ratios significant at .01 level.

[c] $t$ ratios significant at .001 level.

for those classes where the actual evaluations would probably qualify as being reasonably high. In such instances, as for classes A and C in Table 1, significant differences occurred in half of the categories. If evaluations were the only measure for those classes in Tables 2 and 3, the lower expectations, in comparison with still lower--and significant --evaluations would surely escape notice. So, despite the fact that the students in these social science classes exhibited similar directions for contextual expectations and evaluations, the actual intensity of these factors varied. In other words, it now seemed safer to say that extrapolations from evaluations regarding the actual nature of student-teacher classroom interaction could be of questionable value.

The dichotomy between expectations and evaluations appeared more prominent concerning year in college. Here the data substantially indicated that freshmen and sophomores exhibited similar contextual expectations in all three colleges. Regarding evaluations, however, it was the sophomores, not the freshmen who consistently revealed higher evaluations --and, in some cases statistically significant ones as reported in Table 9. This did certainly question the applicability of relating the findings from situational--or generalized--expectations to classroom contexts (Wallace, 1966; Stern, 1970). If a "Freshman Myth" existed for the current study's groupings, t[illegible] it applied to sophomores, too. Somewhat of an inversion took place: compared with sophomores, freshmen expectations appeared more in keeping with their evaluations. As proposed in the conceptual framework outlined earlier, these findings support the need for differentiating between the operation of sets of college-related expectations. In this regard, it might be noted that even in Feldman and Newcomb's comprehensive study, no attempt was made to distinguish between

differing sets of student expectations (1969, I:78-82), such as situational and contextual ones.

Thus, as part of any future research agenda, it would be important to learn how students report each set of expectations. Are the differences to be found in the perception of each set? To be sure, the findings of this study--in keeping with Stogdill's framework--indicated that the role and performance aspects of a classroom would clearly account for some of the possible differences in reporting contextual expectations. In sum, the dynamics of classroom participation compared with institutional membership have still to be clarified.

What implications did this study reveal for currently accepted evaluations? If evaluations simply meant a reaction to the instructional process, then the end-of-course ratings might still serve a valuable function. For, as this study confirmed--in keeping with previous research--an end-of-course rating can be a generally reliable indicator of student reaction. But, if testing student-teacher interaction becomes desirable, then, the appropriateness of expectation inputs would be obvious. Clearly, presently conceived end-of-course ratings have built-in limitations. In addition, these ratings could create the false impression that the students in any one class remain inert, especially if an instructor received similar ratings over time. Certainly for this study's groups, the expectations dimension would have added to an understanding of teacher-student interaction, or lack of it.

The problem of classroom interaction raised a question concerning the contextual expectations model. If this approach inherently called for greater participation by student and teacher in defining classroom objectives, its corollary would be the increased awareness of each other's expectations. Hence, it might be argued that in order to establish the

credibility of this model, the study had to demonstrate the obvious. That is, student expectations would be different from their evaluations. If this was so, it resulted from the fact that end-of-course ratings characteristically failed to include the intensity of student expectations. Obvious or not, the findings of this study would discourage any evaluations cast in isolation from expectations.

## Recommendations

If this study was limited by the in-group homogeneity of the student samples, it might be appropriate to replicate the design in other multi-institutional settings with a cross-section of students. Since most end-of-course ratings research has taken place in state colleges and universities, it would be advantageous to introduce the expectations dimension into those settings, too. If single institutions must be used, then the sample should be heterogeneous. Additional work in the area of instrumentation seemed appropriate. Although this study suggested Osgood's Semantic Differential technique to overcome the increased proliferation of instruments, perhaps other scales--some already in use--could be correlated with it. The thrust of any future refinements, however, should focus on validating expectations-evaluations instruments.

Improvements in the study design could be suggested, too. One would be to introduce certain controls into the design itself. For example, it could now prove worthwhile to manipulate the classroom situation--the independent variable--by employing the concept of contextual expectations into some groups only. Although many factors would require control, it would be important to learn the effects of this approach in actual classrooms. This type of experiment could be conducted in larger institutions with relative ease. In small colleges further controls would have to be introduced through a multi-institutional design.

Further, studies might consider the intervening variables of class size and differing subject areas in relation to the expectations-evaluations approach suggested here. Longer-range longitudinal studies would be needed to determine the impact of actually employing this approach on both students and teachers. Similarly, comparisons between the more innovative curricula and colleges with the more conventional approaches would certainly help determine whether contextual expectations could become an acceptable classroom model. Studies of this type would also help evaluate the systems approaches now being suggested for higher education (Axelrod, 1969).

By way of conclusion, it might be said that despite the rapidly changing college scene, the research scene might still lack reliable knowledge regarding its sine qua non: the learning scene. The situation remained not too different from an observation of ten years ago: ". . . a straight-forward attack on the area of student-instructor role dynamics in the college cultural setting is yet to be made" (Dick, 1961:415). Fortunately, a corrective was recently issued calling for a multi-disciplinary field to study the parameters of classroom dynamics, appropriately named the sociology of learning (Boocock, 1966:1). If its limits were still somewhat undefined, examining the interplay of contextual and situational expectations with related sociological variables in the college setting should be part of its research schedule. The concept of a sociology of learning would thereby give needed direction and shape to classroom evaluations, including the contextual expectations dimension.

REFERENCES

Axelrod, Joseph.
1969 Model Building for Undergraduate Colleges. U.S. Office of Education Contract No. OEC 6-10-106. Washington, D.C.: Educational Resources Information Center.

Berger, Joseph, Thomas Conner and William L. McKeown.
1969 "Evaluations and the Formation and Maintainence of Performance Expectations." Human Relations 22 (December): 481-502.

Boocock, Sarane S.
1966 "Toward a Sociology of Learning: A Selective Review of Existing Literature." Sociology of Education 39 (Winter): 1-45.

Cartwright, Dorwin and Alvin Zander (eds.)
1968 Group Dynamics: Research and Theory. 3d ed. New York: Harper and Row.

Cohen, Arthur M. and Florence B. Brawer.
1969 Measuring Faculty Performance. Washington, D.C.: American Association of Junior Colleges.

Costin, Frank, W. T. Greenough and R. J. Menges.
1971 "Student Ratings of College Teaching: Reliability, Validity, and Usefulness." Review of Educational Research 41 (December): 511-535.

Dick, Harry R.
1961 "Student-Faculty Role Consensus." Southwestern Social Science Quarterly 41 (March): 415-423.

Feldman, Kenneth A. and Theodore M. Newcomb.
1969 The Impact of College on Students. 2 vols. San Francisco: Jossey-Bass, Inc.

Gross, Neal, W. S. Mason and A. W. McEachern.
1958 Explorations in Role Analysis. New York: Wiley.

Isaacson, Robert L., W. J. McKeachie, and J. E. Milholland.
1963 "Correlation of Teacher Personality Variables and Student Ratings." Journal of Educational Psychology 54 (April): 110-117.

Lewis, Lionel S.
1968 "Students' Images of Professors." Educational Forum 32 (January): 185-190.

McKeachie, W. J.
1963 "Research on Teaching at the College and University Level." Pp. 1118-1172 in N. L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally.
1969 "Student Ratings of Faculty." AAUP Bulletin 55 (December): 439-444.

## REFERENCES

McClelland, James N.
1970 "The Effect of Student Evaluations of College Instruction Upon Subsequent Evaluations." California Journal of Educational Research 51 (January): 379-382.

Mueller, John H., K. Schuessler, and H. L. Costner.
1970 Statistical Reasoning in Sociology. 2d ed. New York: Wiley.

Newcomb, Theodore M.
1962 "Student Peer-Croup Influence." Pp. 469-488 in Nevitt Sanford (ed.), The American College. New York: Wiley.

Oetting, S. R.
1967 "The Effect of Forcing Response on the Semantic Differential." Educational and Psychological Measurement 67 (Autumn): 699-702.

Osgood, Charles E., G. J. Suci, and P. H. Tannenbaum.
1957 The Measurement of Meaning. Urbana, Ill.: University of Illinois.

Owen, Paul H.
1967 Some Dimensions of College Teaching: An Exploratory Study Using Critical Incidents and Factor Analysis of Student Ratings. Unpublished Doctoral dissertation. University of Houston.

Perry, Richard R.
1967 "Criteria of Effective Teaching Behavior in an Institution of Higher Education." Proceedings of the Seventh Annual Forum of the Association for Institutional Research. N.p.: The Association for Institutional Research.

Phillips, Beeman N.
1964 "Individual and the Classroom Croup as Frames of Reference in Determining Teacher Effectiveness." Journal of Educational Research 58 (November): 128-131.

Rees, Richard D.
1969 "Dimensions of Students' Points of View in Rating College Teachers." Journal of Educational Psychology 60 (December): 476-482.

Ryans, David G.
1960 Characteristics of Teachers. Washington, D.C.: American Council on Education.

Snider, James G. and Charles E. Osgood (eds.)
1969 Semantic Differential Technique: A Sourcebook. Chicago: Aldine.

Stern, George C.
1970 People in Context: Measuring Person-Environment Congruence in Education and Industry. New York: Wiley.

## REFERENCES

Stogdill, Ralph M.
1959 Individual Behavior and Group Achievement. New York: Oxford.

Turner, Richard L., et al.
1969 "How Do Student Characteristics Affect Their Evaluations of Instructors." Indiana University Educational Bulletin 45 (July): 46-97.

Twa, R. James.
1970 Student and Instructor Expectations of Community College Instructors. Eugene, Ore.: Oregon School Study Council.

Wallace, Walter L.
1966 Student Culture: Social Structure in a Liberal Arts College. Chicago: Aldine.